

NO-A176 222

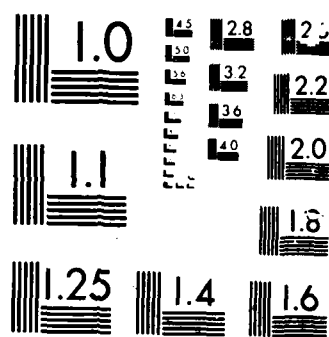
ANALYSIS OF A SIMPLE DEBUGGING MODEL(U) WASHINGTON UNIV 1/1
SEATTLE DEPT OF STATISTICS A E RAFTERY 30 NOV 86 TR-80
N00014-84-C-0169

UNCLASSIFIED

F/G 9/2

NL

11
2. 27
1.00



AD-A176 222

ANALYSIS OF A SIMPLE DEBUGGING MODEL

by

Adrian E. Raftery

TECHNICAL REPORT No. 80

March 1986

Department of Statistics, GN-22

University of Washington

Seattle, Washington 98195 USA



Analysis of a Simple Debugging Model

Adrian E. Raftery

Department of Statistics,
Trinity College,
Dublin 2,
Ireland.

ABSTRACT

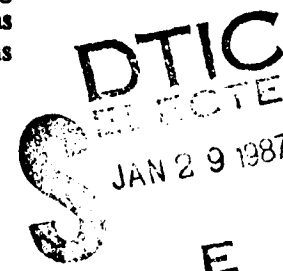
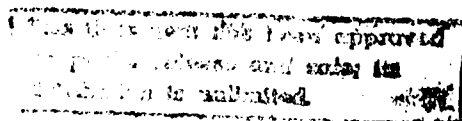
A system has an unknown number of faults. Each fault causes a failure of the system, and is then located and removed. The failure times are independent exponential random variables with common mean. A Bayesian analysis of this model is presented, with emphasis on the situation where vague prior knowledge is represented by limiting, improper, prior forms. This provides a test for reliability growth, estimates of the number of faults, an evaluation of current system reliability, and a prediction of the time to full debugging. Three examples are given.

KEY WORDS: Bayes factor; Improper prior; Non-homogeneous Poisson process; Reliability growth; Software reliability.



Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

Adrian E. Raftery is Lecturer in Statistics, Trinity College, Dublin 2, Ireland, and Visiting Associate Professor of Statistics, University of Washington, Seattle, WA 98195. This work was supported by the Office of Naval Research under contract N00014-84-C-0169. The author is grateful to Fritz Scholz and Peter Guttorp for very helpful comments.



1. INTRODUCTION

Consider a system with an unknown number of faults N . Each fault causes a failure of the system, and is then located and removed. The times at which the N failures occur are assumed to be independent exponential random variables with common mean β^{-1} . Early analyses of this model were carried out by Bazovsky (1961, chap.8) and Cozzolino (1968). It has been much studied in the software reliability literature, where it is often attributed to Jelinski and Moranda (1972).

Problems of interest include finding the probability that all the faults have been removed, estimating the number of remaining faults, evaluating the current reliability of the system, and predicting the time to full debugging. Another question is whether the system's failure rate is decreasing, as the model predicts. Littlewood and Verrall (1981) and Ascher and Feingold (1984, pp.110-111) emphasised the need to test this assumption, and described software reliability data sets in which the failure rate increased over long periods of time.

My aim here is to develop methods which can provide solutions to such problems, as well as a framework for making decisions, such as when to stop debugging. My approach is Bayesian, with an emphasis on the situation where vague prior knowledge about the model parameters is represented by limiting, improper, prior forms.

Much previous research has focussed on point estimation of N (Blumenthal and Marcus 1975; Joe and Reid 1985; Watson and Blumenthal 1980). This is a difficult problem; for example, the *maximum likelihood estimator (MLE)* of N can be infinite with substantial probability. Indeed, Goudie and Goldie (1981), who studied the case where the observed number of failures is specified in advance, concluded that all standard non-Bayesian techniques are liable

to fail. My approach does yield estimators of N ; these are described and compared with other estimators in Section 3.

Forman and Singpurwalla (1977) proposed a stopping rule for debugging the system based on how close the observed likelihood is to a large-sample approximation; their aim was to ascertain whether the system had been fully debugged. Their data are reanalyzed in Section 5. I hope that this paper provides a more precise answer to that question, as well as the basis for a more general stopping rule, which explicitly takes into account the costs associated with the various possible outcomes.

2. TESTING FOR RELIABILITY GROWTH

I assume that the system has been observed for the period $[0, T]$, during which n failures have occurred at times $t = (t_1, \dots, t_n)$, where $n > 1$. I consider the problem of comparing the model described in Section 1 with the constant rate Poisson process $M_0: \lambda(s) = \mu$, where $\lambda(s)$ is the rate of occurrence of failures at time s .

I assume that the sample space consists of systems, rather than of replications of the debugging process for the same system. N is thus a random variable, and I assume that it has a Poisson distribution. It then follows that the model is equivalent to a non-homogeneous Poisson process with rate function

$$M_1: \lambda(s) = \rho \exp(-\beta s) \quad (2.1)$$

where $\rho > 0$ (Scholz 1986). Non-Bayesian statistical analysis of this process has been considered

by Cox and Lewis (1966), Lewis (1972), MacLean (1974), and Berman (1981).

The comparison of M_0 with M_1 is based on the *Bayes factor*, or ratio of posterior to prior odds for M_0 against M_1 ,

$$B_{01} = p(t | M_0) / p(t | M_1) \quad (2.2)$$

the ratio of the marginal likelihoods. In (2.2)

$$p(t | M_0) = \int_0^{\infty} p(t | \mu, M_0) p(\mu | M_0) d\mu$$

$$p(t | M_1) = \int_0^{\infty} \int_0^{\infty} p(t | \rho, \beta, M_1) p(\rho, \beta | M_1) d\rho d\beta$$

If the priors $p(\mu | M_0)$ and $p(\rho, \beta | M_1)$ are proper, (2.2) can be evaluated directly.

I now develop an expression for B_{01} in the situation where vague prior knowledge is represented by limiting, improper, prior forms. I use the standard vague prior for μ

$$p(\mu | M_0) = c_0 \mu^{-1} \quad (2.3)$$

(Jaynes 1968). The likelihood for M_1 is

$$p(t | \rho, \beta, M_1) = \rho^n \exp\{-\beta S - \rho \beta^{-1}(1 - \exp(-\beta T))\}$$

where $S = \sum_{i=1}^n t_i$. This is an exponential family likelihood, for which a natural family of conjugate

prior densities is

$$p(\rho, \beta | M_1) \propto \rho^{k_1} \exp\{-k_2 \beta - k_3 \rho \beta^{-1}(1 - \exp(-\beta T))\} \quad (2.4)$$

Akman and Raftery (1986b) have shown that the unique prior of the form (2.4) for which B_{01} is invariant to scale changes in the time variable and independent of the stopping time T is

$$p(\rho, \beta | M_1) = c_1 \rho^{-2} \quad (2.5)$$

However, the Bayes factor calculated using the improper priors (2.3) and (2.5) involves an arbitrary, undefined, multiplicative constant c_0/c_1 . Akman and Raftery (1986b) have shown how this may be assigned using the minimal imaginary experiment idea of Spiegelhalter and Smith (1982). This consists of imagining that an experiment is performed which yields the smallest possible data set permitting a comparison of M_0 and M_1 , and provides maximum possible support for M_0 . It is then argued that the resulting Bayes factor should be only slightly greater than one. Raftery and Akman (1986) have applied this approach to the change-point Poisson process; their results may be compared with the non-Bayesian solution of Akman and Raftery (1986a).

For the present problem, this procedure yields $c_0/c_1 = \pi^2/6 - 1 = 0.6449$, and

$$B_{01} = 0.6449 (n-1) \left[\int_0^{\infty} \exp(-Ry) \{y/(1-\exp(-y))\}^{n-1} dy \right]^{-1} \quad (2.6)$$

where $R = S/T$. Strictly speaking, any value of B_{01} greater than one indicates that the data provide evidence against reliability growth. However, as a rough order of magnitude interpretation, Jeffreys (1961, Appendix B) has suggested that the evidence should be regarded as strong only if $B_{01} > 10^1$, and as decisive only if $B_{01} > 10^2$.

3. ESTIMATING THE NUMBER OF FAULTS IN THE SYSTEM

The framework developed in Section 2 is used. The results in this section and the next one are conditional on M_1 . If the priors are proper, standard Bayesian inference is straightforward (Akman 1985; Jewell 1985; Langberg and Singpurwalla 1985; Meinhold and Singpurwalla 1983).

It follows from (2.5) that

$$p(N, \beta) = \int_0^{\infty} p(N | \rho, \beta) p(\rho, \beta) d\rho$$

$$\propto (N(N-1))^{-1} \beta^{-1} \quad (3.1)$$

Also,

$$p(t | N, \beta) = \beta^n \exp\{-\beta T(R+N-n)\} N!/(N-n)! \quad (3.2)$$

Combining (3.1) with (3.2), and integrating over β yields the posterior distribution of the number of remaining faults $M=N-n$,

$$p(M | t) \propto (M+R)^{-n} \prod_{i=1}^{n-2} (M+i) \quad (3.3)$$

The probability that the system has been fully debugged is simply $P[M=0|t]$. Interval estimates of N , such as *highest posterior density (HPD)* regions, or one-sided intervals, may readily be found from (3.3).

In many applications, estimation of N is an intermediate step in the solution of other problems. However, if a point estimator of N is required, it may be obtained from (3.3) by

combining it with an appropriate loss function. The posterior mode, \hat{N}_{mod} , is the estimator which corresponds to a zero-one loss function, so that, if the appropriate loss function is bounded, \hat{N}_{mod} may well be a good approximation. The posterior median, \hat{N}_{med} (found by linear interpolation), is an estimator which corresponds to one unbounded loss function, and is also a useful summary of the posterior distribution.

Other point estimators of N which are always finite include Blumenthal and Marcus (1975)'s modified maximum likelihood estimator N^* , and Joe and Reid (1985)'s harmonic mean estimator \tilde{N} . Watson and Blumenthal (1980) considered three other estimators, but their performance in a simulation study was very similar to that of N^* , so I do not consider them further here.

The four estimators, \hat{N}_{mod} , \hat{N}_{med} , N^* , and \tilde{N} , were compared in a small simulation study whose results are summarised in Table 1. β was fixed at 1.0, and T was set equal to $-\log(1-Q)$, where N and Q were fixed at the values shown. Q is thus the probability of a randomly chosen bug causing the system to fail before time T . The results are conditional on $n > 1$.

Table 1 about here

The most striking feature of Table 1 is how badly all four estimators performed; none did much better than an estimator which is identically equal to n . Also, no one estimator was uniformly better than any other. For $Q=0.9$, corresponding to the situation where the system is close to being fully debugged, \hat{N}_{mod} performed best, while for $Q=0.25$, \tilde{N} performed best. These results suggest that it would be better to report the full posterior distribution (3.3), or some

of its salient characteristics, than any one point estimator.

4. ESTIMATING SYSTEM RELIABILITY AND TIME TO FINAL DEBUGGING

The reliability of the system is the probability that it operates without failure for a further, specified, period of length x , say. This is equal to $P[X > x | t]$, where $X = t_{n+1} - T$ is the time to the next failure. Now, $X = \infty$ if $M = 0$, and $P[X > x | M, \beta] = \exp(-M\beta x)$ ($M \geq 1$), so that

$$\begin{aligned} P[X > x | t] &= P[M = 0 | t] + \sum_{M=1}^{\infty} \int_0^{\infty} \exp(-M\beta x) p(M, \beta | t) d\beta \\ &= P[M = 0 | t] + \sum_{M=1}^{\infty} p(M | t) \{1 + M(M+R)^{-1}(x/T)\}^{-n} \end{aligned}$$

$E[X | t]$ is always infinite, but we can calculate

$$E[X | t, M \geq 1] = T(n-1)^{-1} \{1 + R(1 - P[M = 0 | t])^{-1} \sum_{M=1}^{\infty} M^{-1} p(M | t)\}$$

The time to final debugging of the system is $Z = t_N - T$. $Z = 0$ if $M = 0$, while if $M \geq 1$, Z is the maximum of M independent exponential random variables with mean β^{-1} . Thus

$$P[Z \leq z | t] = P[M = 0 | t] + \sum_{M=1}^{\infty} p(M | t) \sum_{k=0}^M (-1)^k \binom{M}{k} \{1 + kz/T(M+R)\}^{-n}$$

and

$$E[Z | t] = \sum_{M=1}^{\infty} p(M | t)(M+R) \sum_{k=1}^M (-1)^{k+1} \binom{M}{k} k^{-1}$$

5. EXAMPLES

I now apply the techniques proposed here, as well as those of Blumenthal and Marcus (1975) and Joe and Reid (1985), to several, previously analyzed, software reliability data sets. The results are given in Table 2.

Table 2 about here

Example 1. Goel and Okumoto (1979) gave the failure times of a piece of software developed as part of the Naval Tactical Data System. These data had previously been analyzed by Jelinski and Moranda (1972). By the end of the production and testing phases, which lasted 540 days, 31 failures had occurred.

The Bayes factor B_{01} , at about 10^{-3} , indicated decisive evidence for reliability growth, but $P[M=0|\epsilon]$ was only 0.27, indicating that the system had probably not been fully debugged. Indeed, three further failures later occurred.

The techniques proposed here gave similar results to the likelihood analysis of Joe and Reid (1985). \hat{M}_{med} and \bar{M} were very close. The 0.5 likelihood interval, proposed as an interval estimator by Joe and Reid (1985), had coverage probability close to 0.76, and was the same as the 76% HPD region based on (3.3).

Example 2. Meinhold and Singpurwalla (1983) gave the failure times of a real-time command and control system. These data have also been analyzed by Musa (1975), Goel (1985), and Okumoto (1985). After $n=7$ failures, the MLE of N was infinite, and an analysis at this point

revealed differences between the present approach and a likelihood analysis. For example, the 0.5 likelihood interval was $2-\infty$ and had coverage probability less than 0.6, but posterior probability 0.88 from (3.3).

Meinhold and Singpurwalla (1983) suggested a Bayesian analysis with a proper prior for N which was Poisson with mean 50. This yielded a posterior distribution for M concentrated between 21 and 57; by comparison (3.3) yielded the 95% HPD region 0-155. In fact, 129 further failures occurred.

After $n=136$ failures, the present approach and the likelihood analysis of Joe and Reid (1985) gave results which were in close agreement, as in Example 1.

Example 3. Forman and Singpurwalla (1977) analyzed a data set consisting of 107 failures. The data were grouped, and, like them, I have assumed that the average time of occurrence within each group was at the center of the time interval.

After $n=8$ failures, the procedure of Joe and Reid (1985) produced an interval estimate for M which included all its possible values, but whose coverage probability was less than 0.64. After $n=99$ failures, the probability of eight or more faults remaining was less than 10^{-4} . Thus, the fact that eight more failures did occur casts doubt on the appropriateness of the model for this data set.

REFERENCES

- Akman, V.E. (1985). "Some Aspects of Statistical Inference for Stochastic Processes," unpublished Ph.D. thesis, Trinity College, Dublin, Department of Statistics.
- Akman, V.E., Raftery, A.E. (1986a), "Asymptotic Inference for a Change Point Process," *Annals of Statistics*, 14, no. 4, to appear.
- Akman, V.E., Raftery, A.E. (1986b), "Bayes Factors for Non-Homogeneous Poisson Processes With Vague Prior Information," submitted to *Journal of the Royal Statistical Society. Ser. B*.
- Ascher, H., Feingold, H. (1984), *Repairable Systems Reliability*, New York: Marcel Dekker.
- Bazovsky, I. (1961), *Reliability Theory and Practice*, New Jersey: Prentice-Hall.
- Berman, M. (1981), "Inhomogeneous and Modulated Gamma Processes," *Biometrika*, 68, 143-152.
- Blumenthal, S., and Marcus, R. (1975), "Estimating Population Size With Exponential Failure," *Journal of the American Statistical Association*, 70, 913-922.
- Cox, D.R., and Lewis, P.A.W. (1966), *The Statistical Analysis of Series of Events*, London: Methuen.
- Cozzolino, J.M. (1968), "Probabilistic Models of Decreasing Failure Rate Processes," *Naval Research Logistics Quarterly*, 15, 361-374.
- Forman, E.H., and Singpurwalla, N.D. (1977), "An Empirical Stopping Rule for Debugging and Testing Computer Software," *Journal of the American Statistical Association*, 72, 750-757.

- Goel, A.L. (1985), "Software Reliability Models: Assumptions, Limitations, and Applicability," *IEEE Transactions on Software Engineering*, SE-11, 1411-1423.
- Goel, A.L., and Okumoto, K. (1979), "Time-Dependent Error-Detection Rate Model for Software Reliability and Other Performance Measures," *IEEE Transactions on Reliability*, R-28, 206-211.
- Goudie, I.B.J., and Goldie, C.M. (1981), "Initial Size Estimation for the Linear Pure Death Process," *Biometrika*, 68, 543-550.
- Jaynes, E.T. (1968), "Prior Probabilities," *IEEE Transactions on Systems Science and Cybernetics*, SSC-4, 227-241.
- Jeffreys, H. (1961), *Theory of Probability* (3rd ed.), Oxford: University Press.
- Jelinski, Z., and Moranda, P.B. (1972), "Software Reliability Research," in *Statistical Computer Performance Evaluation*, ed. W. Freiberger, London: Academic Press, pp.465-484.
- Jewell, W.S. (1985), "Bayesian Extensions to a Basic Model of Software Reliability," *IEEE Transactions on Software Engineering*, SE-12, 1465-1471.
- Joe, H., and Reid, N. (1985), "Estimating the Number of Faults in a System," *Journal of the American Statistical Association*, 80, 222-226.
- Langberg, N., and Singpurwalla, N.D. (1985), "Unification of Some Software Reliability Models via the Bayesian Approach," *SIAM Journal of Scientific and Statistical Computing*, 6, 781-790.
- Lewis, P.A.W. (1972), "Recent Results in the Statistical Analysis of Univariate Point Processes," in *Stochastic Point Processes*, ed. P.A.W. Lewis, New York: Wiley, pp. 1-54.

- Littlewood, B., and Verrall, J.L. (1981), "Likelihood Function of a Debugging Model for Computer Software Reliability," *IEEE Transactions on Reliability*, R-30, 145-148.
- MacLean, C.J. (1974). "Estimation and Testing of an Exponential Polynomial Rate Function Within the Non-Stationary Poisson Process," *Biometrika*, 61, 81-85.
- Meinhold, R.J., and Singpurwalla, N.D. (1983), "Bayesian Analysis of a Commonly Used Model for Describing Software Failures," *The Statistician*, 32, 168-173.
- Musa, J.D. (1975), "A Theory of Software Reliability and its Applications," *IEEE Transactions on Software Engineering*, SE-1, 312-327.
- Okumoto, K. (1985), "A Statistical Method for Software Quality Control," *IEEE Transactions on Software Engineering*, SE-11, 1424-1431.
- Raftery, A.E., and Akman, V.E. (1986), "Bayesian Analysis of a Poisson Process With a Change-Point," *Biometrika*, 73, no.1, to appear.
- Scholz, F.-W. (1986), "Software Reliability Modeling and Analysis," *IEEE Transactions on Software Engineering*, SE-12, 25-31.
- Spiegelhalter, D.J., and Smith, A.F.M. (1982), "Bayes Factors for Linear and Log-Linear Models With Vague Prior Information," *Journal of the Royal Statistical Society, Ser. B*, 44, 377-387.
- Watson, D.A., and Blumenthal, S. (1980), "Estimating the Size of a Truncated Sample," *Communications in Statistics - Theory and Methods*, A9, 1535-1550.

Table 1. Root Mean Squared Error of Point Estimators of N^a

Q	N	Root Mean Squared Error of Estimator			
		\hat{N}_{mod}	\hat{N}_{med}	N^*	\tilde{N}
0.9	10	1.30	2.29	1.37	2.81
0.9	100	5.59	6.16	5.64	5.94
0.25	10	6.97	6.51	6.80	5.19
0.25	100	63.39	46.49	57.68	46.36

^a 200 simulations with $n > 1$ for each value of (Q, N) .

Table 2. Results for Examples 1,2,3.

Example	n	R	$\log_{10} B_{01}$	\hat{M}_{mod}	\hat{M}_{med}	$P[M=0 t]$	95% HPDR	\hat{M}	M^*	\bar{M}	0.5 LI
1	31	8.4	-3.0	0	.9	.27	0-7	0	1	1	0-3
2	7	4.1	.7	2	9.1	.05	0-155	∞	4	10	2- ∞
2	136	38.0	-16.0	6	6.5	.01	1-16	6	6	7	3-11
3	8	4.0	.6	1	6.9	.07	0-125	6	3	8	0- ∞
3	24	8.4	-.2	2	4.3	.07	0-33	2	3	4	0-9
3	99	18.4	-27.2	0	.0	.56	0-2	0	0	1	0-1
3	107	13.3	-45.8	0	.0	.95	0	0	0	0	0

NOTE: $i-j$ denotes the set of integers from i to j inclusive. Notation: $\hat{M}_{mod} = \hat{N}_{mod} - n$; $\hat{M}_{med} = \hat{N}_{med} - n$; \hat{M} is the MLE of M ; $M^* = N^* - n$; $\bar{M} = \bar{N} - n$. 95% HPDR is the 95% HPD region from (3.3), while 0.5 LI is the 0.5 likelihood interval defined by Joe and Reid (1985).

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 80	2. GOVT ACCESSION NO. ADA176222	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Analysis of a simple debugging model.		5. TYPE OF REPORT & PERIOD COVERED TR 12/1/85-11/30/86
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Adreian E. Raftery		8. CONTRACT OR GRANT NUMBER(s) N00014-84-C-0169
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics, C-22 University of Washington Seattle, WA 98195		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-661-003
11. CONTROLLING OFFICE NAME AND ADDRESS ONR Code N63374 1107 NE 45th St. — Seattle, WA 98105		12. REPORT DATE March 1986
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Bayes factor; improper prior; non-homogeneous Poisson process; reliability growth; software reliability.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)		

END

2-87

DTIC